PUSL3123 AI and Machine Learning

20 CREDIT MODULE /

C1: 50% Exam
C2: 50% Group coursework

MODULE LEADER: Dr Neamah Al-Naffakh

MODULE AIMS

- To introduce the area of machine learning (ML) covering unsupervised, supervised and reinforcement learning from Bayesian perspectives.
- To review and apply learning techniques and artificial intelligence concepts towards data representations and decision-making and prediction systems.
- To enable students to analyse real datasets and control real-time systems.

ASSESSED LEARNING OUTCOMES (ALO):

1. Describe and analyse a range of artificial intelligence methods and their applications.
2. Apply the concepts of unsupervised, supervised and reinforcement learning to tackle machine learning problems.
3. Demonstrate the ability to implement and apply machine learning techniques to make decisions on artificial and real data sets

OVERVIEW

This document contains all the necessary information pertaining to the assessment of PUSL3123 *AI and Machine Learning*. The module is assessed via 50% exam and 50**% Group coursework**.

The sections that follow will detail the assessment tasks that are to be undertaken. The submission and expected feedback dates are presented in Table 1. All assessments are to be submitted electronically via the respective DLE module pages before the stated deadlines.

Table 1 : Assessment Deadlines

	Submission Deadline	Feedback
Group Coursework (50%)	3rd January 2024	Within 20 working days
Exam (50%)	25th January 2024	Immediately

All assessments will be introduced in class to provide further clarity over what is expected and how you can access support and formative feedback prior to submission. Whilst the assessment information is provided at the start of the module, it is not necessarily expected you will start this immediately – as you will often not have sufficient understanding of the topic. The module leader will provide guidance in this respect.

Task 1– Literature Review (40%)

This task contributes to **40%** of the overall group coursework mark for PUSL3123. The focus of this task to undertake a review on ***Pattern Recognition*** by identifying at least **NINE** research papers in the area of **biometrics** (e.g., face, fingerprint, iris, voice, gait, keystroke, behavioural profiling, activity recognition- **select one topic only**).

The marking scheme will be structured as follows:

You are required to write a report to cover any selected topic in the area of biometric pattern recogntion. The individual report should be no more than **3,000** words (excluding diagrams, images, tables, and references). The report should be organised as follows:

- **Introduction and conclusion (5 Marks)** – are the selected topic introduced/ described appropriately? Your introduction should clearly explain the overall research topic and the depth of the information to be presented. For the conclusion section, do mention the limitations of the selected study and their implications if not already addressed in the discussion section of the paper. DON'T introduce new information. Instead, your conclusion should restate your main points and provide closure. **(750 words)**

- **Literature review (20 Marks)** – are the critical analysis done properly? For example, but not limited to, data collection, feature extraction/selection (if any), classifier/s, and system evaluation (i.e., system performance). **(1500 words)**

- **Discussion (10 Marks)** - how does each of the above parameters has an impact on the proposed system. You should describe, analyze, and interpret the findings of the selected research papers. **(750 words)**

- **Report format and Language (5 Marks)-** Is the report well-presented and structured? Are the references sufficient and properly cited in the report?

Task 2: Implementation of Three Machine Learning Algorithms (60%)

This assignment contributes to **60%** of the overall group coursework mark for PUSL3123. The focus of this task is to assess your understanding of machine learning techniques. You are required to write MATLAB code for implementing the following algorithms.

1- Neural network
2- K Means
3- K-Nearest Neighbors (KNN)

Task 2.1 – Data statistics (5 Marks)

Data Set Information:

The Iris data set is widely used as a beginner's dataset for machine learning purposes. The data set contains 3 classes of 50 instances each, where each class refers to a type of Iris plant (i.e., a total of 150 records under five attributes - sepal length, sepal width, petal length, petal width and species).

A. To get data for supervised Neural network, load the MATLAB file named ***fisheriris.mat.***

B. Report N - the total number of rows (objects, cases) in your data.

C. For each column (feature) from 1 to 4 **report:** the Mean, the Standard Deviation, Maximum, Minimum and Root Mean Square.

Task 2.2 – Neural Network Setup (20 Marks)

1. Shuffle the dataset randomly by selecting 60% for training and remaining 40% for testing. Once data is randomly selected, create training data, training target, testing data and testing target.

2. Select Feedforward Recognition Neural Networks (feedforwardnet)

3. Define the hidden layer size for NN using iteration (assume **10, 15, and 20** hidden layers) and for each neural network setting (i.e., for each hidden layer) repeat the experiment 3-4 times.

4. Train the neural network based on the training dataset created and view the trained net (similar as Figure 1, input and output numbers are just for illustration).

5. Test the trained net using testing dataset and evaluate its performance (e.g. classifier accuracy). You may run several times for training and testing to get an average performance value (i.e., average percentage of correct classifications).

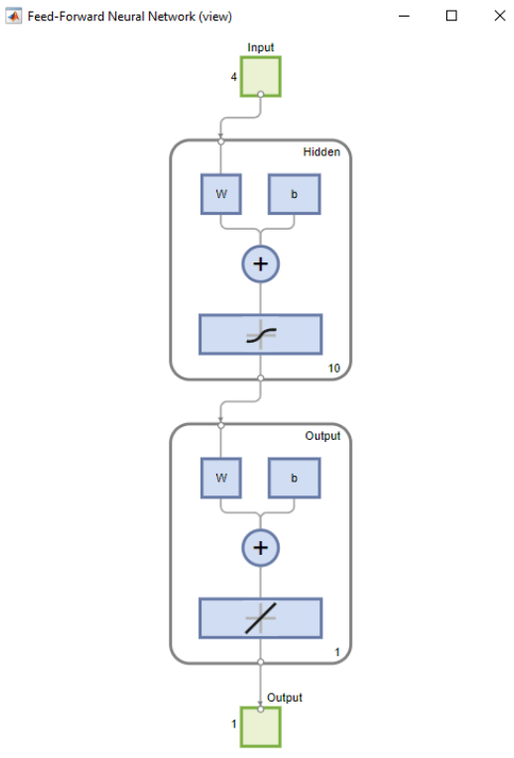

Figure 1: Feed-Forward NN (view)

Task 2.3 – K Means (20 Marks)

1. To get data for your KMean clustering, load the MATLAB file named `kmeansdata.mat`
2. Given the number of clusters as 3 (i.e., K=3), implement Kmeans clustering and then repeat the same procedure (i.e., use iteration) to evaluate different number of clusters (i.e., K=3, 4, and 5) to find out the optimal number of classes that achieve the best performance.
3. For each K value, report the mean performance using the Silhouette measure and plot the Silhouette for each cluster (each K value) as shown in Figure 2.
4. What are the stopping criteria for Kmeans clustering? Plot the clusters and the cluster centroids, which should be similar as in Figure 2 (Note: Figure 3 is an example).
5. Report (via MATLAB code) the best number of clusters and explain why?
6. From your observation and analysis, what are the limitations or drawbacks of Kmeans clustering?

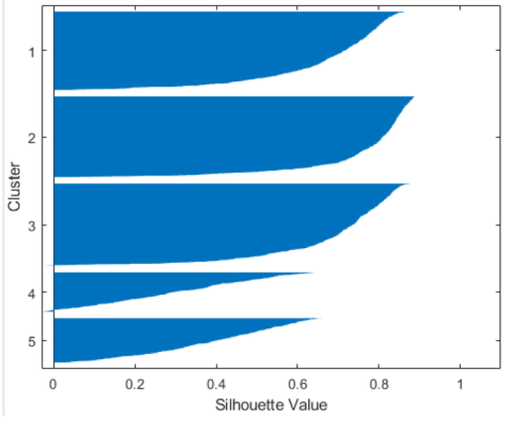

Figure 2: Silhouette measure for 5 clusters

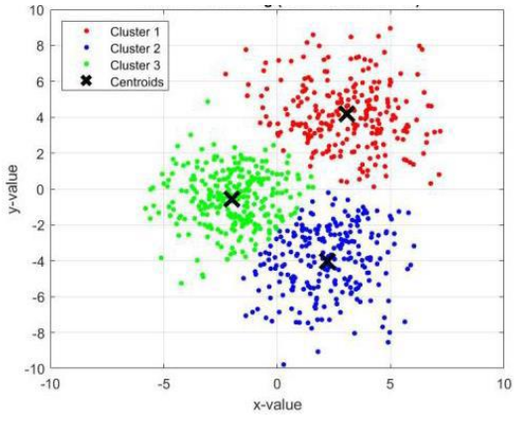

Figure 3: After Kmeans clustering

Task 2.4 – Implementation of KNN (15 Marks)

1. To get data for your supervised classifiers (K-NN), load the MATLAB file named ***fisheriris.mat***
2. Shuffle the dataset randomly by selecting 60% for training and remaining 40% for testing. Once data is randomly selected, create training data, training target, testing data and testing target.
3. Evaluate different K values (i.e., K=5 and 7) by utilizing MATLAB functions *fitcknn.*
4. For each K value, report the confusion matrix and percentage of correct classifications
5. From your observation and analysis, what are the limitations or drawbacks of KNN?

- **Note-** use iteration/s when required to avoid repeating certain statements in your script, for example, use loop for evaluating different K values and to reporting the confusion matrix and percentage of correct classification rates.

	Data preparation (3M)	**set up Feedforward NN (2M)**	**Define hidden layer size (5)**	**Train and test the NN (5M)**	**Evaluate the performance (5M)**
NN Implementation, (20)	Shuffle the dataset into training and testing The efficiency and organisation of the code is checked	Set up the NN using Feed-Forward	Using nested loop to define the hidden layer size and run the experiment 3-4 times. The efficiency and organisation of the code is checked	Train and test the NN classifier A proper documentation of the MATLAB script	Calculate the Average correct classification rate for each hidden layer

	Implementation of Kmeans (4)	**Evaluation of K means (4)**	**Plot (4)**	**Report the optimal K an explain why (4)**	**drawbacks oof Kmeans (4)**
K means implementation, (20%)	Implement KMeans clustering to evaluate different number of clusters (i.e., K=3, 4, and 5) by using iteration Ensure the code is efficient.	For each K value, report the mean performance using the Silhouette measure and plot the Silhouette for each K value	Plot the clusters and the cluster centroids Highlight the stopping criteria for Kmeans clustering	Report the best number of clusters and explain why?	From your observation and analysis, what are the limitations or drawbacks of KMeans clustering

	Data preparation (2)	**Evaluate KNN (3)**	**Report KNN Result (8)**	**Drawbacks oof KNN (2)**
KNN implementation, (15%)	**Load the dataset** Ensure the code is efficient.	Evaluating different K values (i.e., K=5, 7) by using iteration Ensure the code is efficient.	For each K value, report the confusion matrix and percentage of correct classifications. Ensure the code is efficient.	From your observation and analysis, what are the limitations or drawbacks of KNN

Any references should be appropriately cited in the report. Harvard referencing style is recommended. You should write your report as concisely as possible, and it is important that you do not exceed or within 10% of the allowed word limit.

It is important that all members of the group contribute towards all sections of the report. It is NOT appropriate for one member to complete the background section and another group member to complete the evaluation.

In terms of marking, all group members will receive the same marks. However, should a group member not actively contribute to the creation of the report, please ensure you email me (neamah.al-naffakh@plymouth.ac.uk) and **all members of the group** (to ensure transparency) and I will note this against your group. Upon assessing the submission, I will ensure the report is marked on an individual basis.

Submission Guidelines

- Both the Literature Report and MATLAB code must be submitted to the DLE (**Please submit Task 1 & 2 as a single PDF on the DLE**) by the specified submission date.
- Ensure to attach the code in the Appendix (screenshots of your code is **NOT allowed** and would not be considered).
- If you have any queries on submission or in relation to the referred work, please contact the Module Leader in the first instance, if they are unavailable please contact the Faculty Office on 01752 584584 immediately OR nsbmplympartnership@plymouth.ac.uk so any problems can be rectified.
- PLEASE NOTE that we cannot accept work submitted via email.

General Guidance

Extenuating Circumstances

There may be a time during this module where you experience a serious situation which has a significant impact on your ability to complete the assessments. The definition of these can be found in the University Policy on Extenuating Circumstances here:

https://www.plymouth.ac.uk/uploads/production/document/path/15/15317/Extenuating_Circumstances_Policy_and_Procedures.pdf

Plagiarism

All of your work must be of your own words. You must use references for your sources, however you acquire them. Where you wish to use quotations, these must be a very minor part of your overall work. To copy another person's work is viewed as plagiarism and is not allowed. Any issues of plagiarism and any form of academic dishonesty are treated very seriously. All your work must be your own and other sources must be identified as being theirs, not yours. The copying of another persons' work could result in a penalty being invoked.

Further information on plagiarism policy can be found here:
Plagiarism: https://www.plymouth.ac.uk/student-life/your-studies/essential-information/regulations/plagiarism

Examination Offences:
https://www.plymouth.ac.uk/student-life/your-studies/essential-information/exams/exam-rules-and-regulations/examination-offences

Turnitin (http://www.turnitinuk.com/) is an Internet-based 'originality checking tool' which allows documents to be compared with content on the Internet, in journals and in an archive of previously submitted works. It can help to detect unintentional or deliberate plagiarism.
It is a formative tool that makes it easy for students to review their citations and referencing as an aid to learning good academic practice. Turnitin produces an 'originality report' to help guide you. To learn more about Turnitin go to:
https://guides.turnitin.com/01_Manuals_and_Guides/Student/Student_User_Manual

Referencing

The University of Plymouth Library has produced an online support referencing guide which is available here: http://plymouth.libguides.com/referencing
Another recommended referencing resource is Cite Them Right Online; this is an online resource which provides you with specific guidance about how to reference lots of different types of materials.
The Learn Higher Network has also provided a number of documents to support students with referencing: